

畳み込みニューラルネットワーク Convolutional Neural Networks (Part 1)

東京大学 大学院情報理工学系研究科
創造情報学専攻 講師
中山 英樹

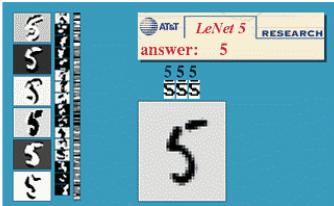
MACHINE PERCEPTION GROUP

目次

- ▶ 1. 画像認識イントロダクション
- ▶ 2. 置み込みニューラルネットワーク
- ▶ 3. 演習課題

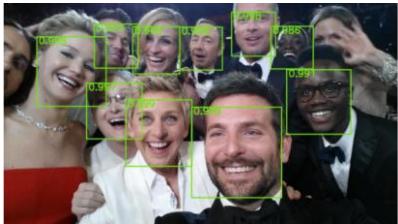


画像認識タスクの例



文字認識

<http://yann.lecun.com/exdb/lenet/index.html>
<http://arxiv.org/pdf/1412.1842v1.pdf>



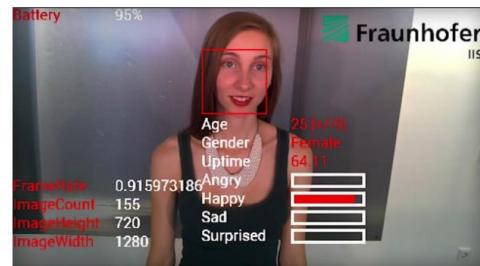
顔検出

<https://www.technologyreview.com/s/535201/the-face-detection-algorithm-set-to-revolutionize-image-search/>



人検出

<http://www.nextplatform.com/2015/08/10/google-research-boasts-deep-learning-detection-engine-with-gpus/>



表情認識

<http://www.extremetech.com/extreme/189259-real-time-emotion-detection-with-google-glass-an-awesome-creepy-taste-of-the-future-of-wearable-computers>



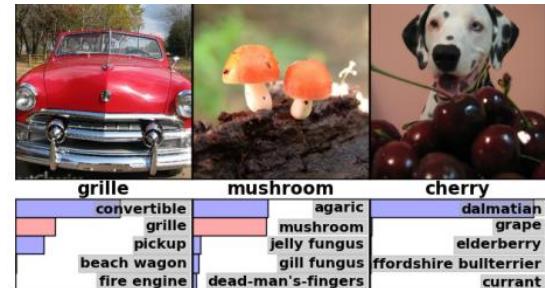
姿勢推定

<http://static.googleusercontent.com/media/research.google.com/ja/pubs/archive/42237.pdf>



特定物体認識

<http://www.csie.ntu.edu.tw/~winston/projects/teldap.html>



一般物体認識

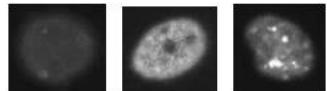
<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

アプリケーション

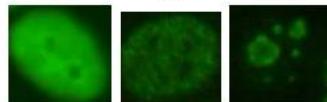


自動運転

<http://deepdriving.cs.princeton.edu/>



Homogeneous 2494 Speckled 2831 Nucleolar 2598



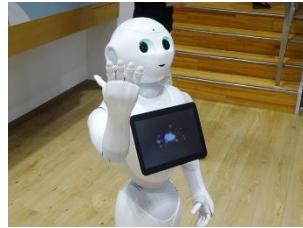
Homogeneous 150 Speckled 109 Nucleolar 102

医療

<http://arxiv.org/pdf/1504.02531v2.pdf>

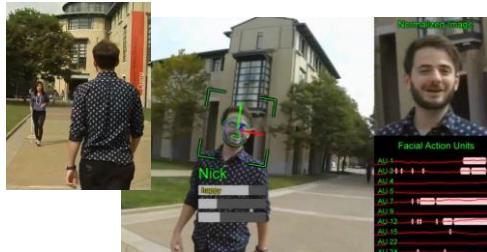
ファッショ

http://www.lv-nus.org/papers/2012/magic_closet-MM12.pdf



ロボット

<http://3media.biz/wadai/softbank-ginza-robot-papper.html>



視覚障碍者支援

<http://www.cmu.edu/news/stories/archives/2015/october/blind-navigation-app.html>



介護・見守り

<http://www.rbbtoday.com/article/2015/10/08/135914.html>



セキュリティ

<http://www.itmedia.co.jp/news/articles/1511/11/news146.html>



環境保護

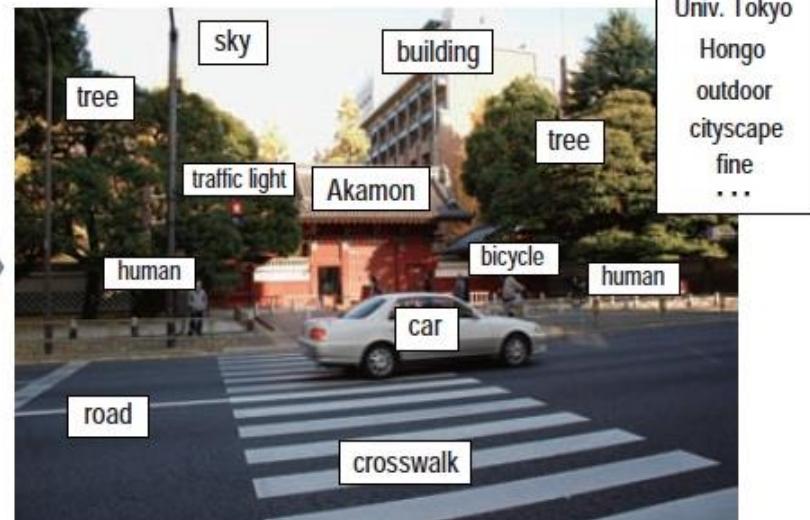
<http://leafsnap.com/>

一般物体認識（一般画像認識）

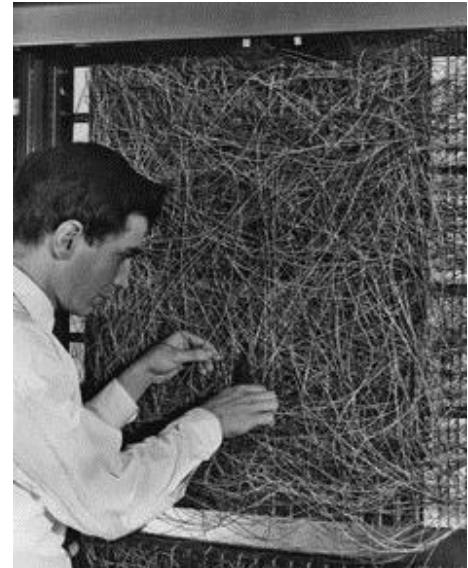
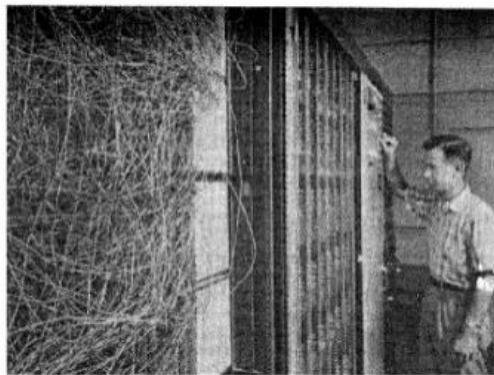
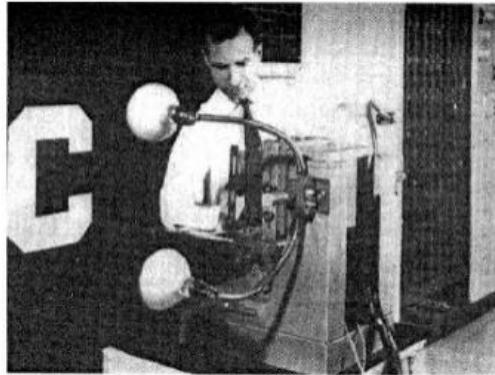
▶ 制約をおかない実世界環境の画像を言語で記述

- 一般的な物体やシーン、形容詞、印象語
- 2000年代以降急速に発展（コンピュータビジョンの人気分野）
- 幅広い応用先

デジタルカメラ、ウェアラブルデバイス、画像検索、ロボット、…



Rosenblatt's perceptron (1960)

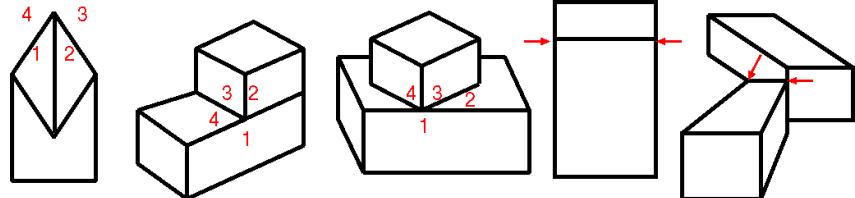


- ▶ 400個の光センサ
- ▶ 結合重みはポテンショメータ(可変抵抗)

歴史：モデルベースドから統計的学习へ

▶ 1970年代

- 線画解釈(積み木の世界)



▶ 1980年代

- 幾何形状モデルベース (identification)
- 一般化円筒
- 3次元モデル

▶ 1990年代

- アピアランスベースド, 統計的手法に基づくアプローチ
- Eigenface, 固有空間法

▶ 2000年代

- 局所特徴, 機械学習の進歩によるブーム到来

アピアランスベースド

- ▶ 二次元入力のさまざまな事例から本質的な(=不变な)特徴を学習

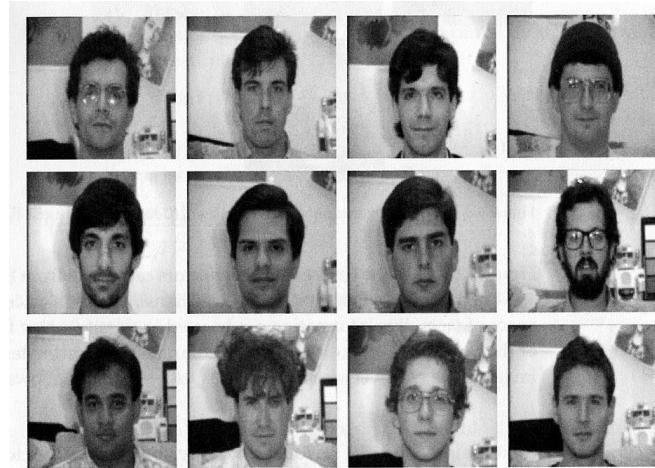


[Yan et al. ICCV2007]

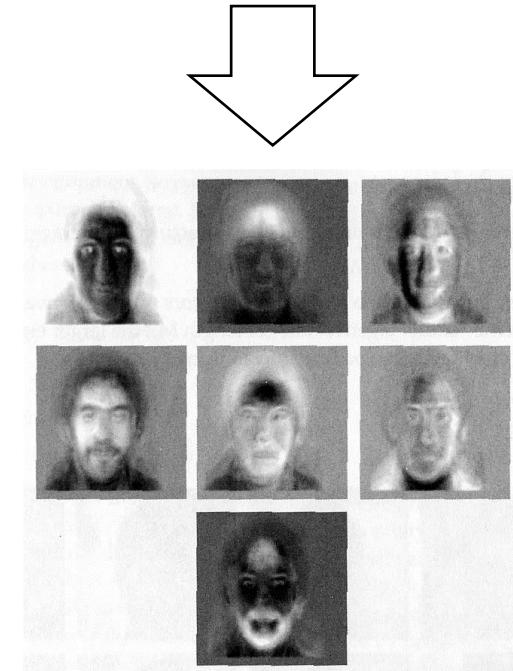
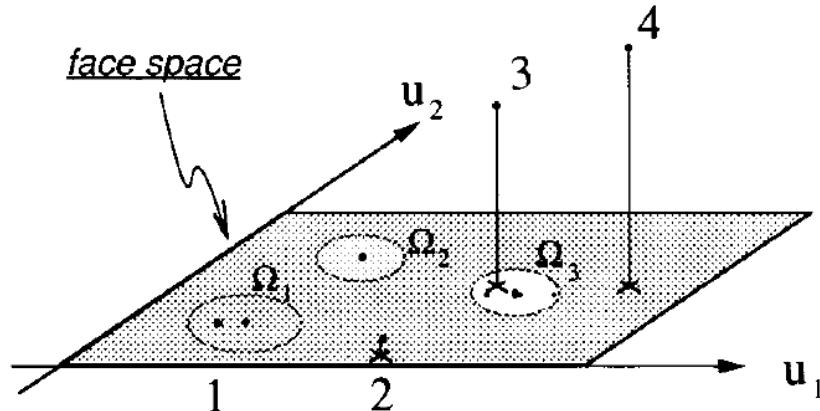
Eigenfaces

[Turk & Pentland, CVPR1991]

- ▶ 画像はベクトル1個(画素値)
- ▶ 学習サンプルから固有空間を求める

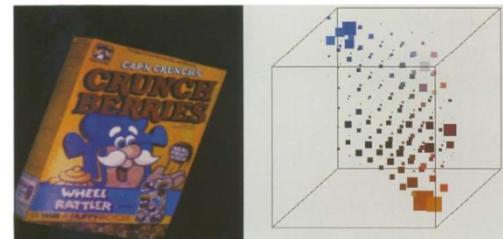


- ▶ 固有ベクトル=固有顔
- ▶ 入力画像を固有空間に投影、検索



Color indexing [Swain and Ballard, IJCV1991]

- ▶ カラーヒストグラムを用いた画像検索



セマンティック・ギャップ

▶ 事例の“類似度”をどう定義すべきか？

- 例えば、単純なカラーヒストグラム（色の割合）だと右の二つの画像は非常に近い値となる

I look my dog contest:
<http://www.hemmy.net/2006/06/25/i-look-like-my-dog-contest/>



- ## ▶ もともと物理的な信号に過ぎない画像と“意味”との間には大きな隔たりがある

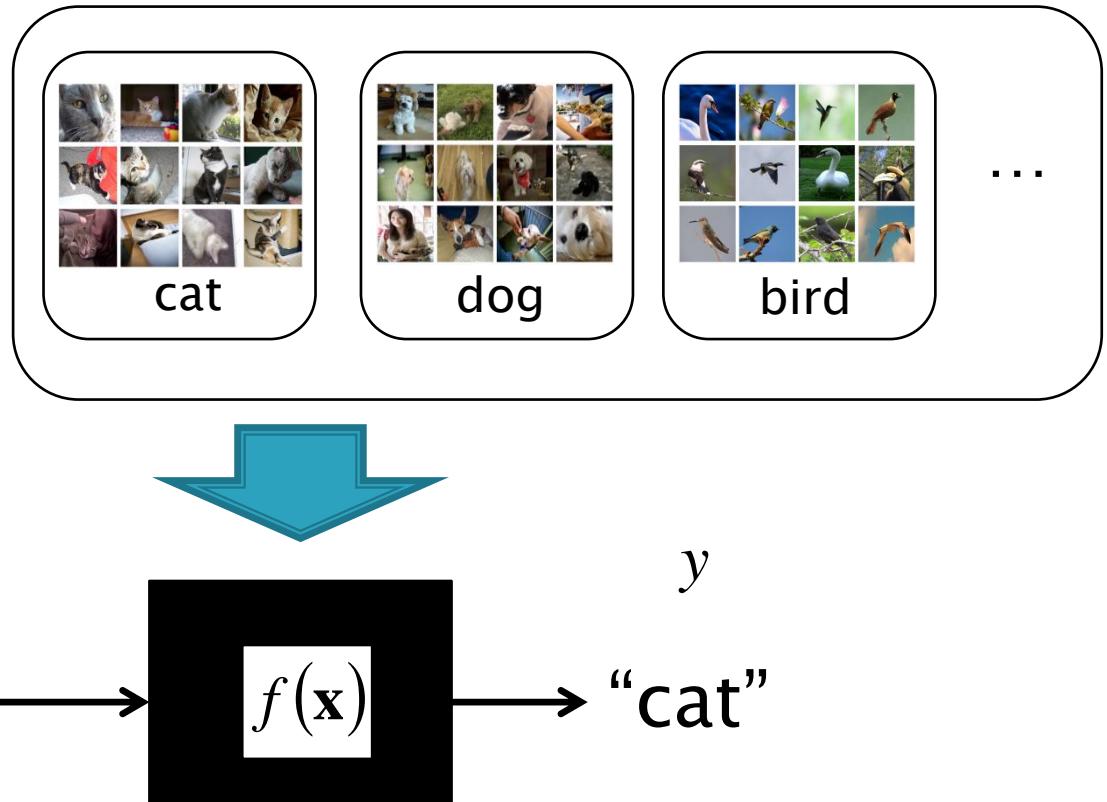
- ## ▶ どういう視覚的手掛けり(特徴)を見ればよいか？

特徴を自動的に見つけさせる

▶ 機械学習(教師付)

$$\{(\mathbf{x}_i, y_i), i = 1, \dots, N\}$$

大量のラベル付き訓練データ
(x:画像, y:ラベル)



未知のデータ(学習データに含まれない)を正しく認識させることが目標

場合の数？



チェス 10^{120}



将棋 10^{220}



囲碁 10^{360}



32×32サイズのカラー画像(8bit)

10^{7400}

Torralba et al., “80 million tiny images: a large dataset for non-parametric object and scene recognition”, TPAMI, 2008.



Figure from
[Ramanan et al, ICCV'09]

Large-scale recognition



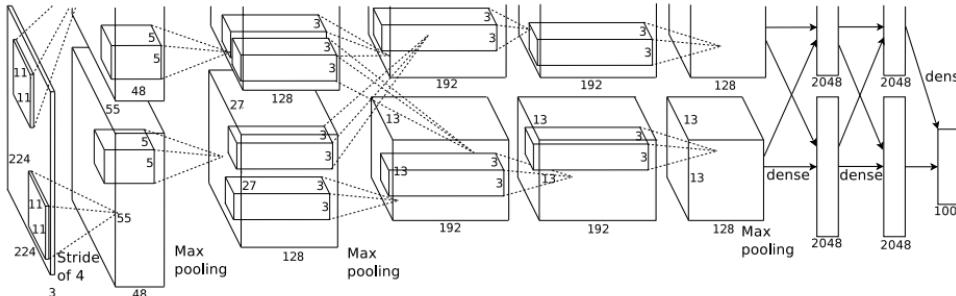
2010
カテゴリ数: $10^3 \sim 10^4$
サンプル数: $10^6 \sim 10^7$



Figure from
Russakovsky et al.,
ILSVRC'14 slides.

ILSVRC 2012 でのブレークスルー

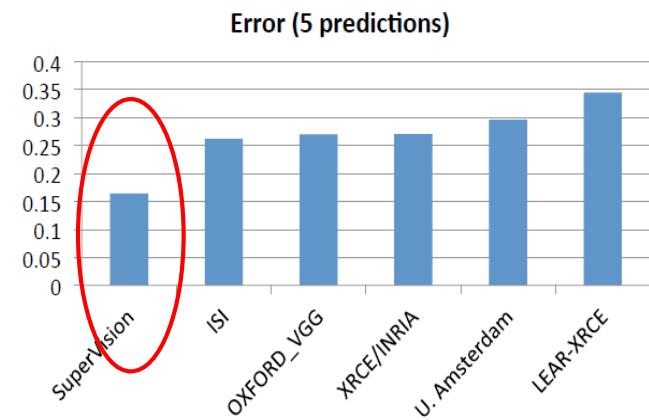
- 1000クラス識別タスクで、deep learning を用いたシステムが圧勝
 - トロント大学Hinton先生のチーム (AlexNet)



[A. Krizhevsky *et al.*, NIPS'12]

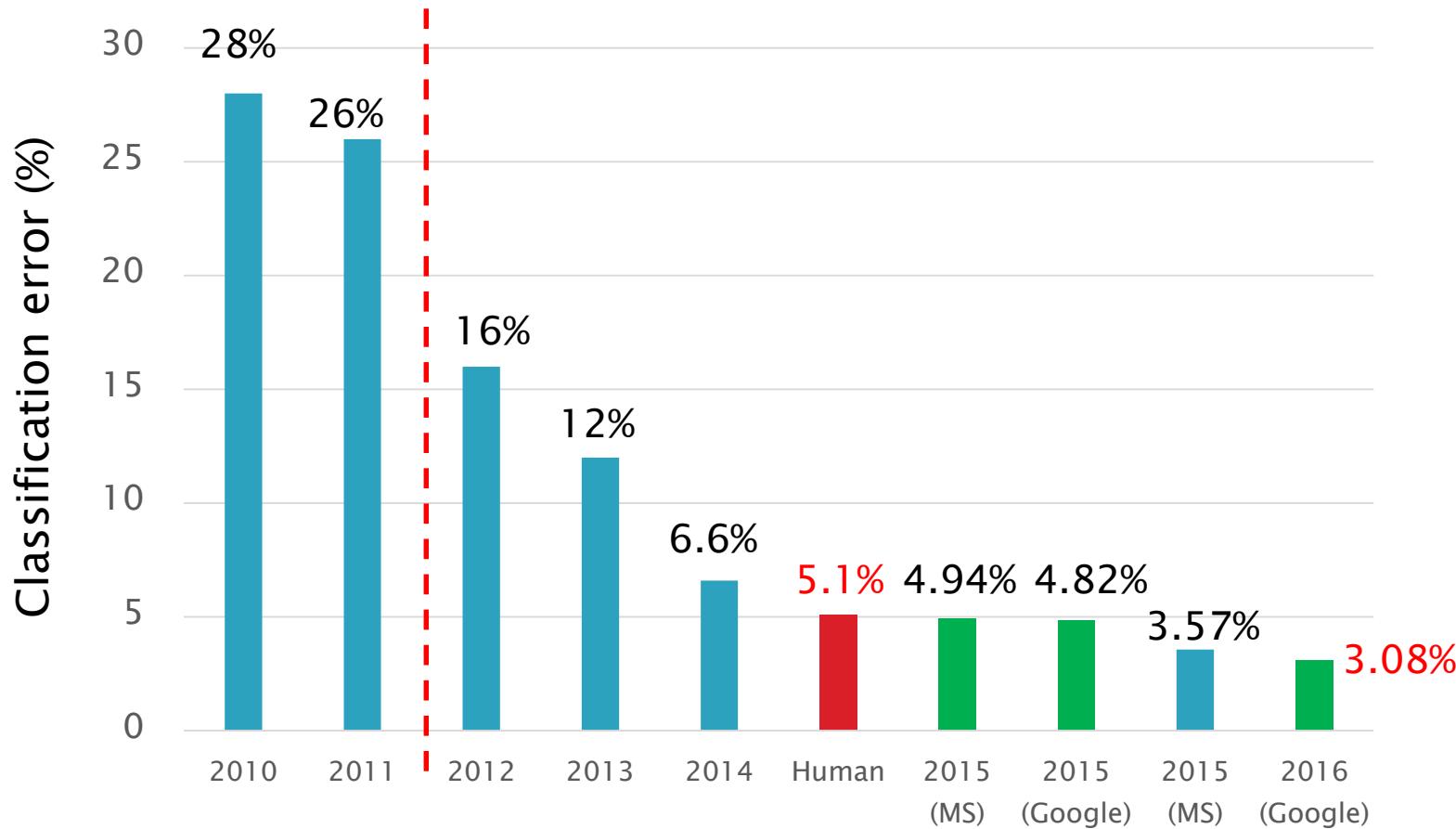


エラー率が一気に10%以上減少！
(※過去数年間での向上は1~2%)



圧倒的な性能向上

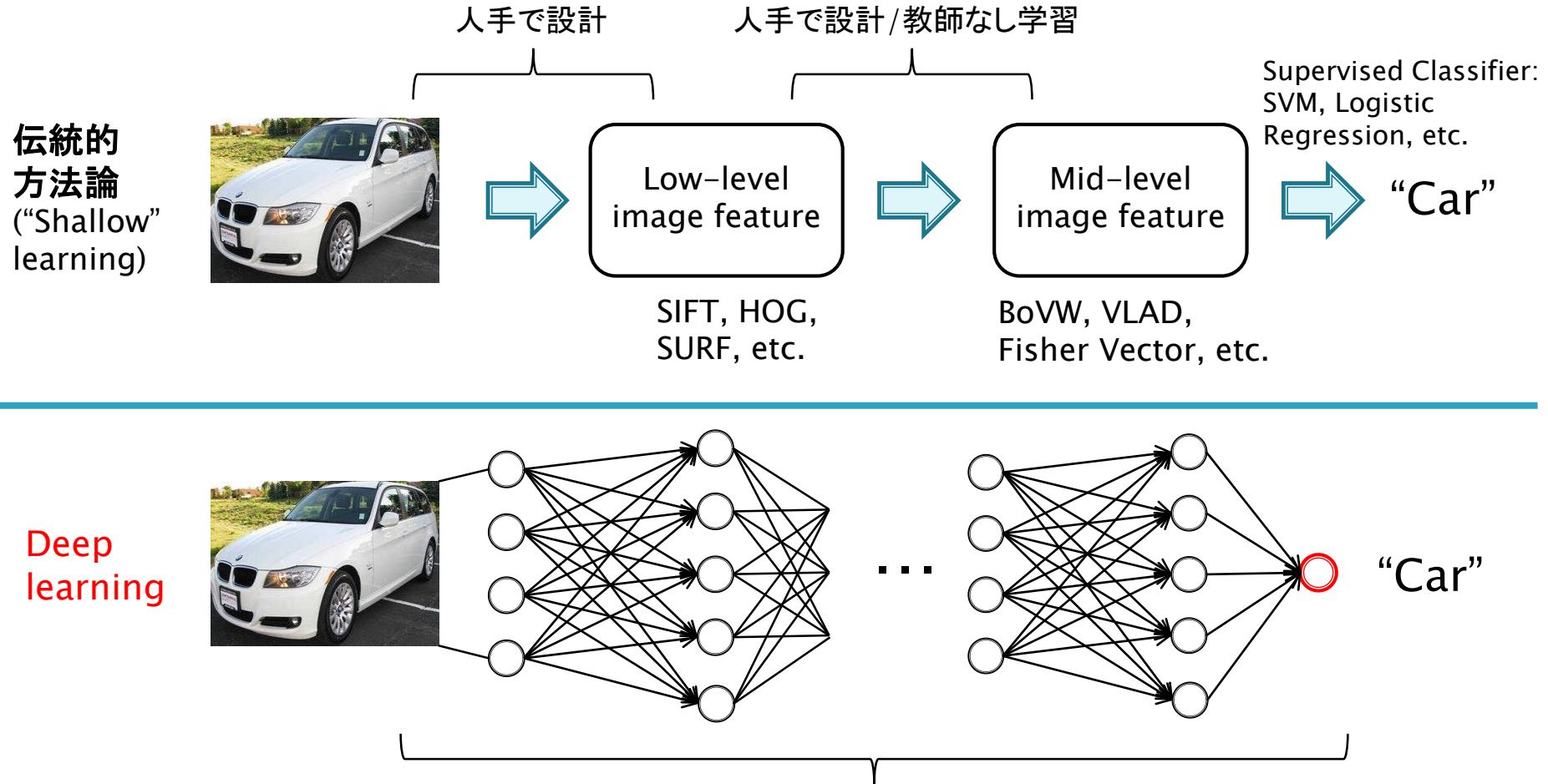
- ▶ エラー率が 16% (2012) → 3.08% (2015)



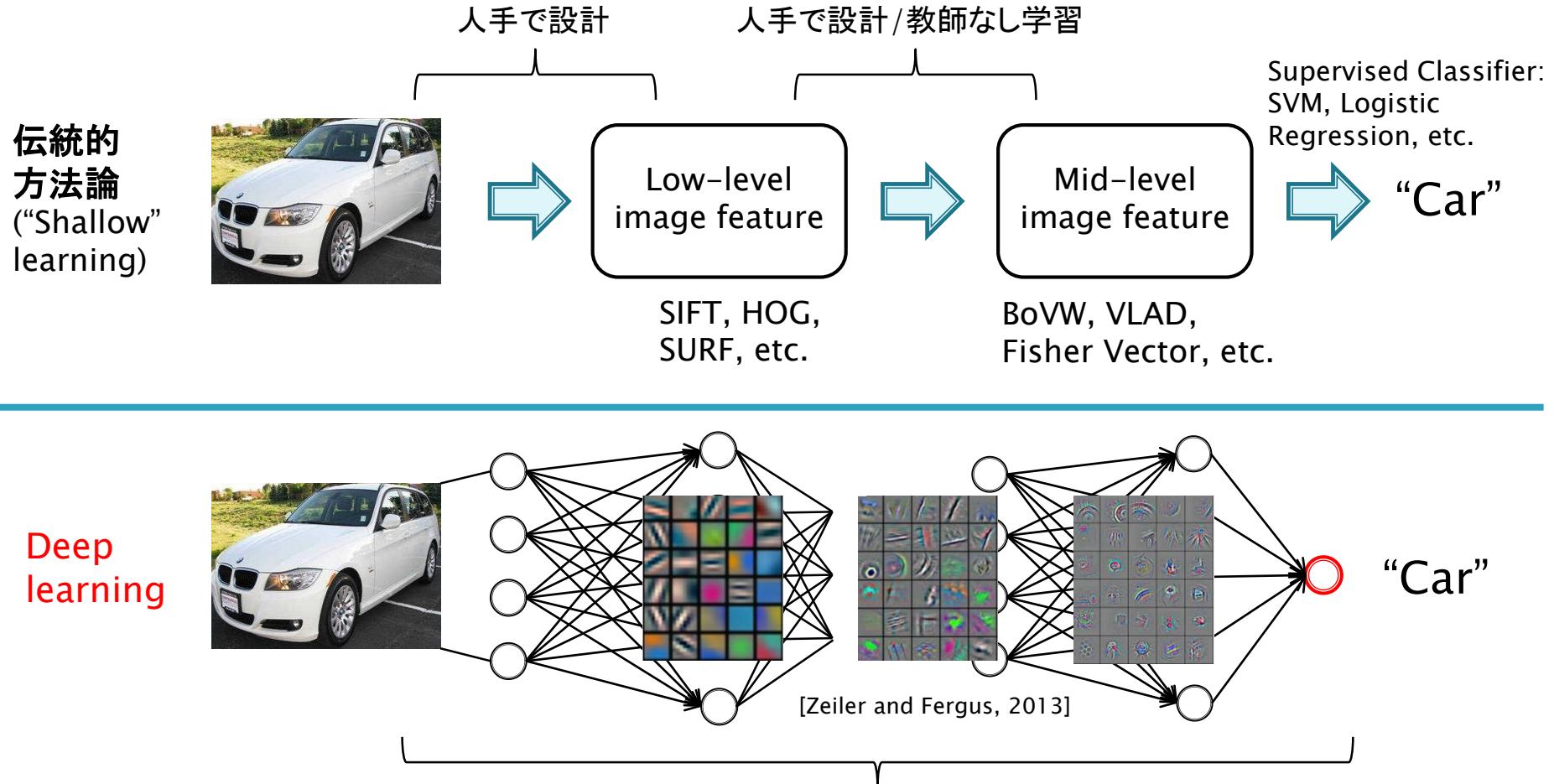
He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", arXiv, 2015.

Szegedy et al., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", arXiv, 2016.

画像認識パイプラインの変化



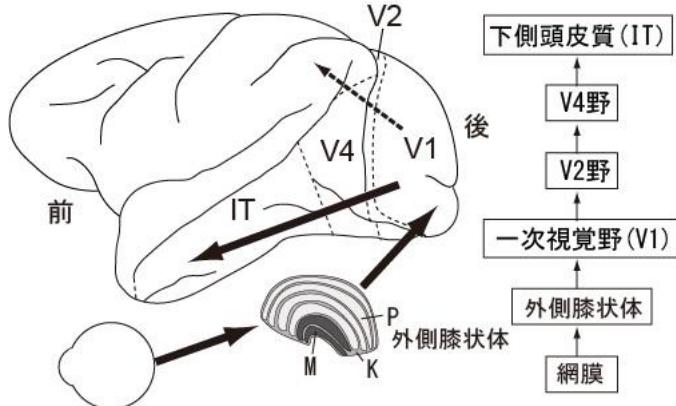
画像認識パイプラインの変化



目次

- ▶ 1. 画像認識イントロダクション
- ▶ 2. 置み込みニューラルネットワーク
- ▶ 3. 演習課題

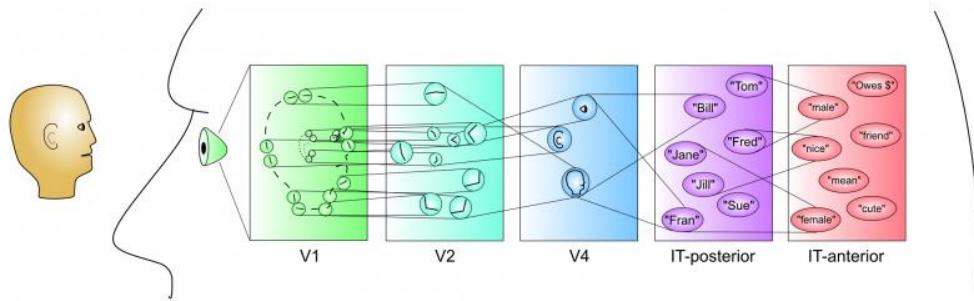
脳(視覚野)の生理学的知見



マカクザルの視覚経路

<https://bsd.neuroinf.jp/wiki/色選択性細胞>

- ▶ 階層構造を有する
- ▶ 段階的に抽象的な特徴に反応するニューロンが現れる
- ▶ 網膜からV1までは、網膜像の位相幾何学的構造が保たれるように投射される
- ▶ V1ニューロンは受容野を持つ

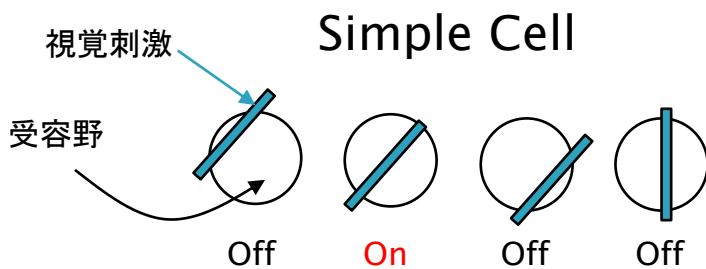
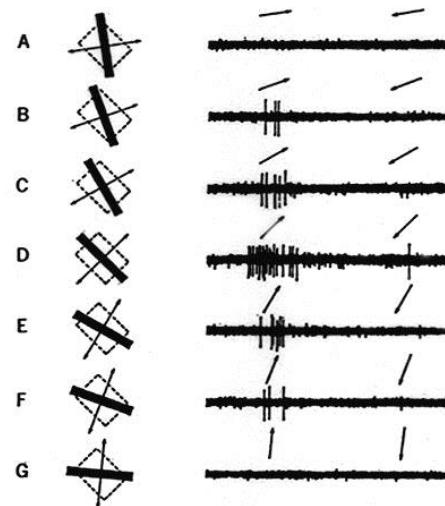
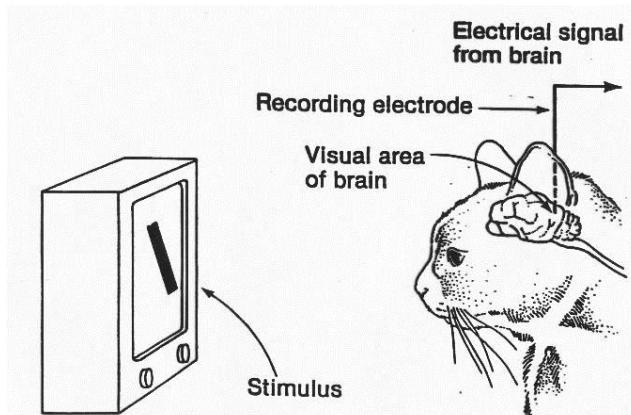


<https://grey.colorado.edu/CompCogNeuro/index.php/CCNBook/Perception>

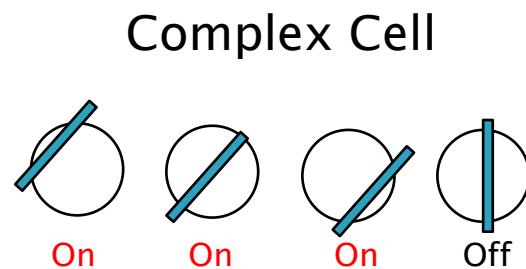
一次視覚野(V1)

▶ Hubel & Wiesel, 1959

- 猫の一次視覚野の各ニューロンが、入力パターン(エッジ)の向きに対して選択的に反応することを発見 (1981年ノーベル賞)



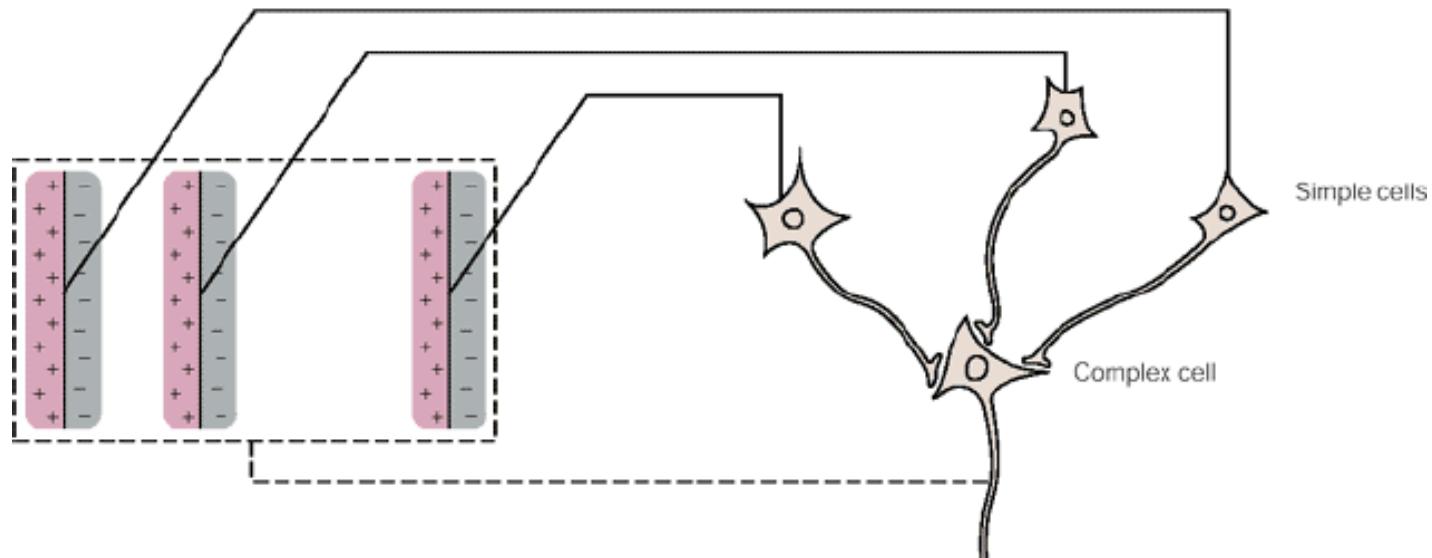
位置・方向に対する選択性



方向のみに対する選択性
(位置に対する不変性)

Hubel-Wiesel 階層仮説

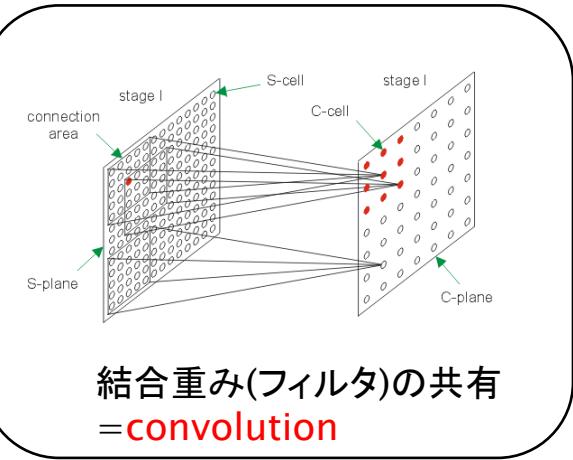
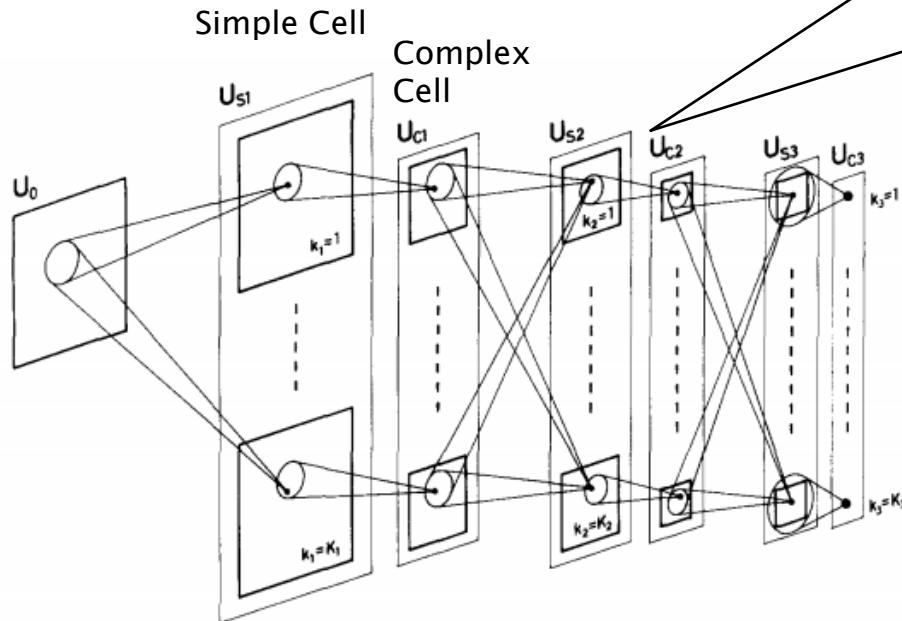
- Simple cell の出力を合成すればComplex cell の挙動は説明できる（という仮説）



<https://kin450-neurophysiology.wikispaces.com/Visual+Cortical+Neurons>

Neocognitron

- ▶ 福島邦彦先生、1980年代前後
 - 畳み込みニューラルネットワークの原型
 - Simple Cell → Complex Cell の結合はpoolingに対応
 - 段階的に解像度を落としながら、局所的な相関パターンを抽出

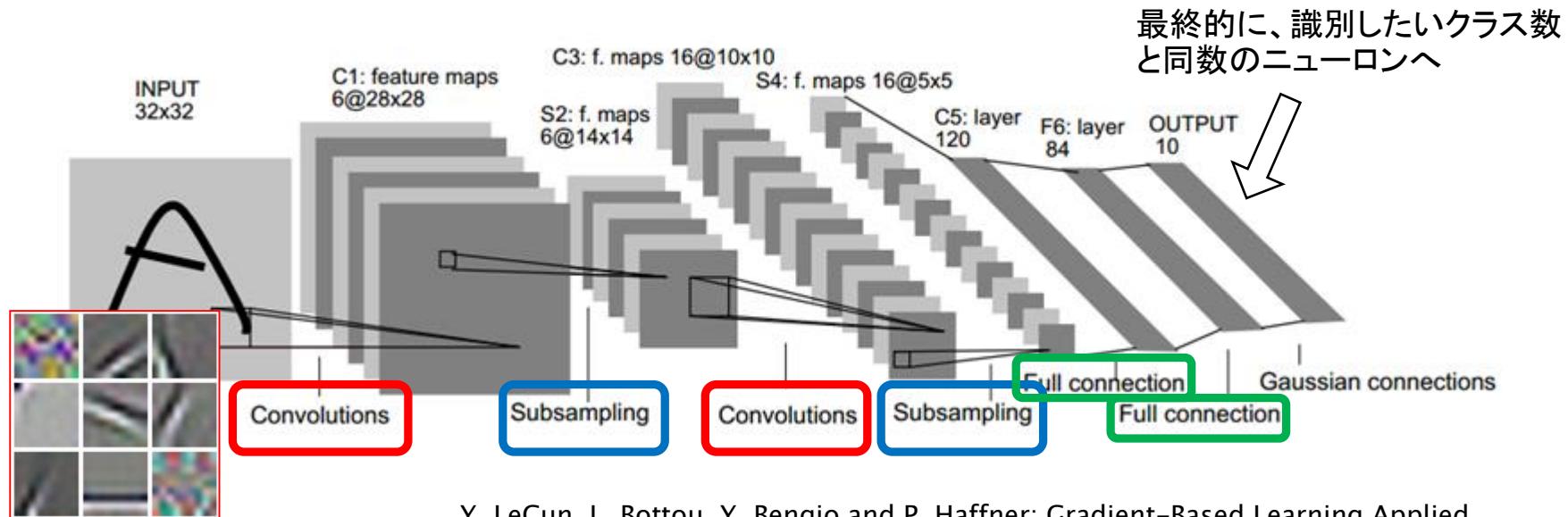


<http://www.kiv.zcu.cz/studies/predmety/uir/NS/Neocognitron/en/func-C-cell.html>

Kunihiko Fukushima, "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biological Cybernetics, 36(4): 93–202, 1980.

Convolutional neural network (CNN)

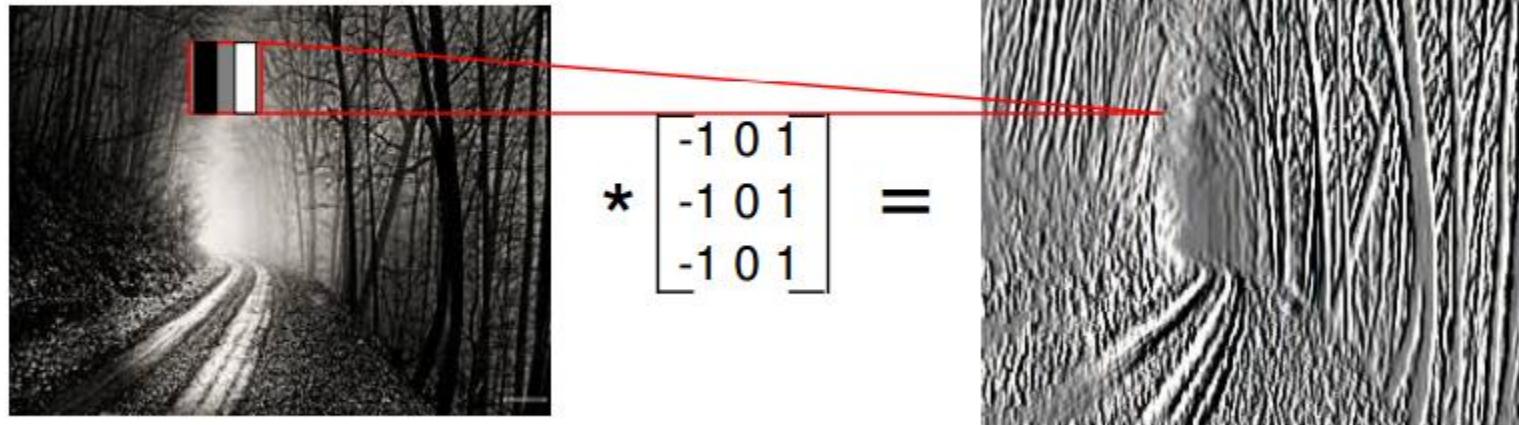
- ▶ 局所領域(受容野)の畳み込みとプーリングを繰り返す多層ペーセプトロン
 - 単純な全結合ネットワークと比べて大幅にパラメータ数を削減
 - 入力の位置に関する不变性
 - 誤差逆伝播法による全体最適化



Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278–2324, 1998.

Convolution

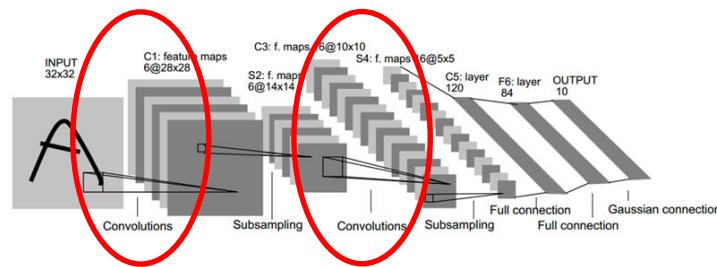
- ▶ 一般的なフィルタだと…
 - 例) エッジ抽出



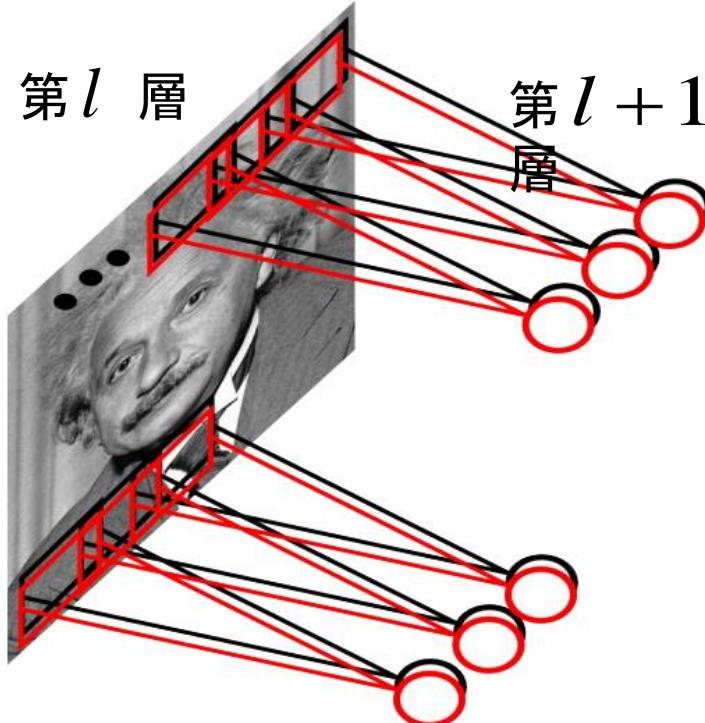
- ▶ CNNでは識別に有効なフィルタ係数をデータから学習する

Source: M. Ranzato, CVPR'14 tutorial slides

畳み込み層



- ▶ 各フィルタのパラメータは全ての場所で共有
 - 色の違いは異なる畳み込みフィルタを示す



※もちろん入力は生画像のみ
とは限らない(中間層など)

非線形活性化関数(とても重要)

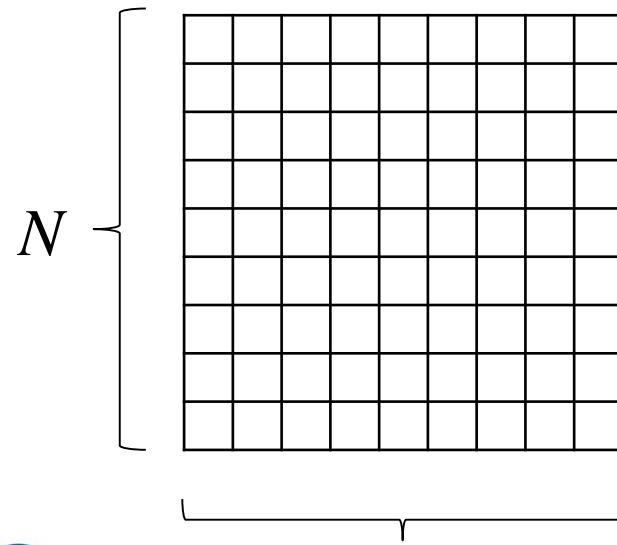
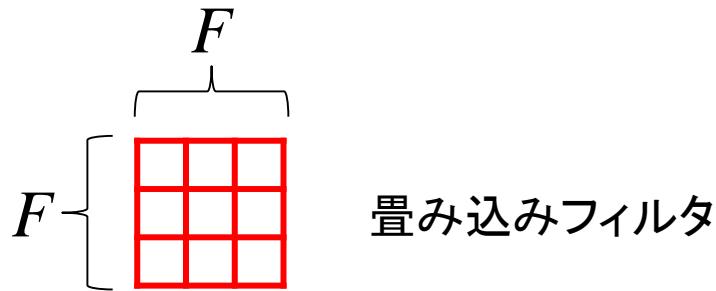
$$\mathbf{z}^{l+1} = h(\mathbf{W}^{l+1} * \mathbf{z}^l + \mathbf{b}^{l+1})$$

フイルタの係数 入力 バイアス

Source: M. Ranzato, CVPR'14 tutorial slides

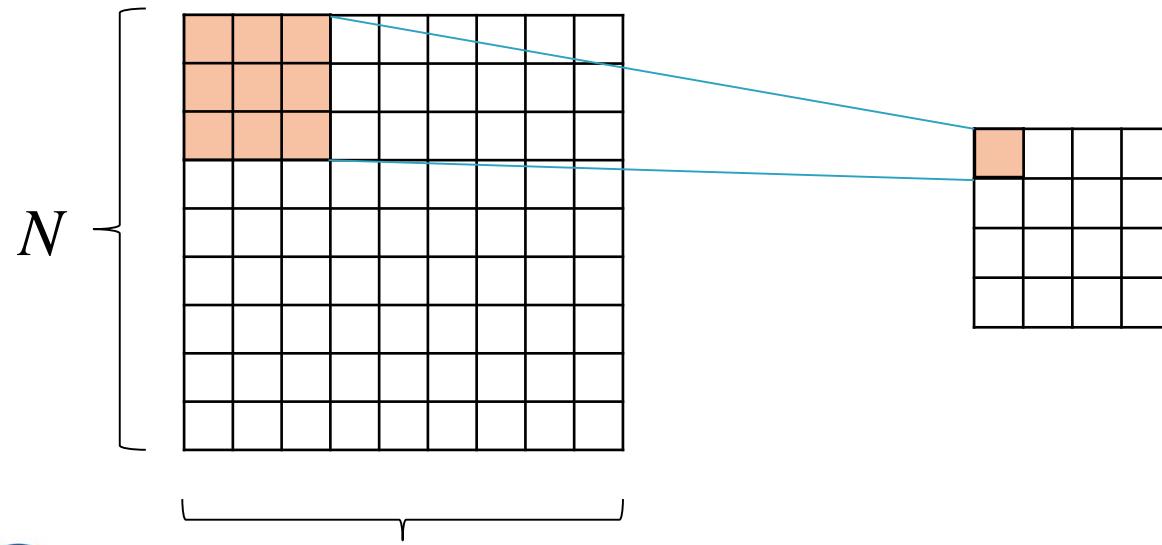
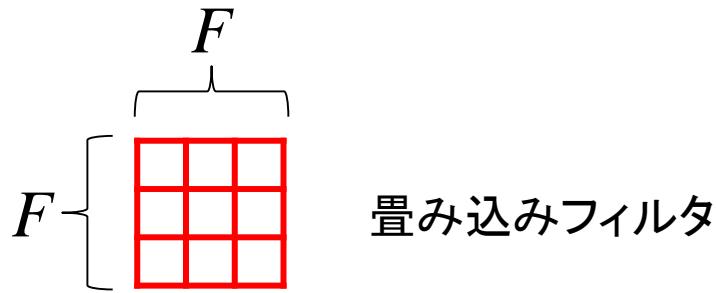
畳み込み層

▶ (まず簡単のため) 入力一層、フィルタ一つの場合



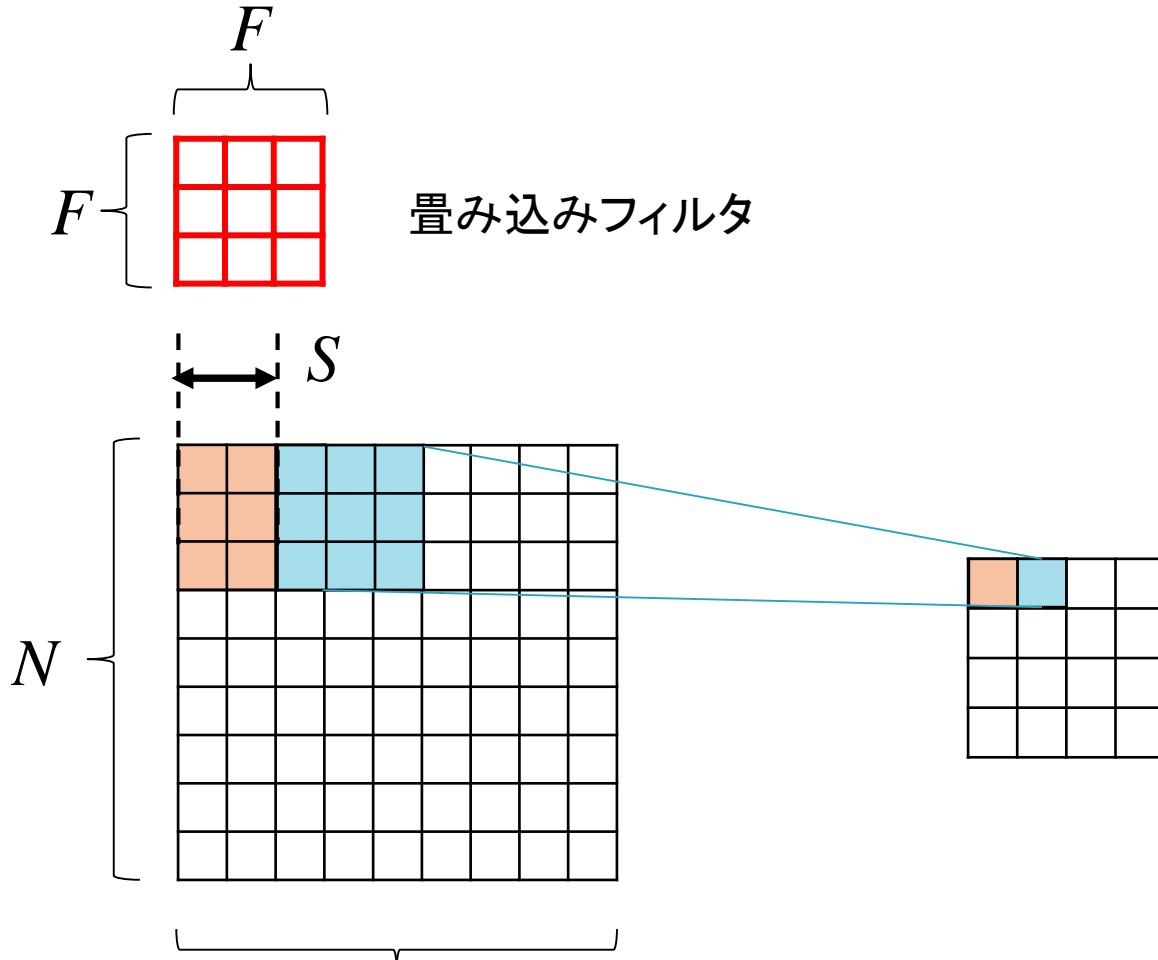
畳み込み層

▶ (まず簡単のため) 入力一層、フィルタ一つの場合



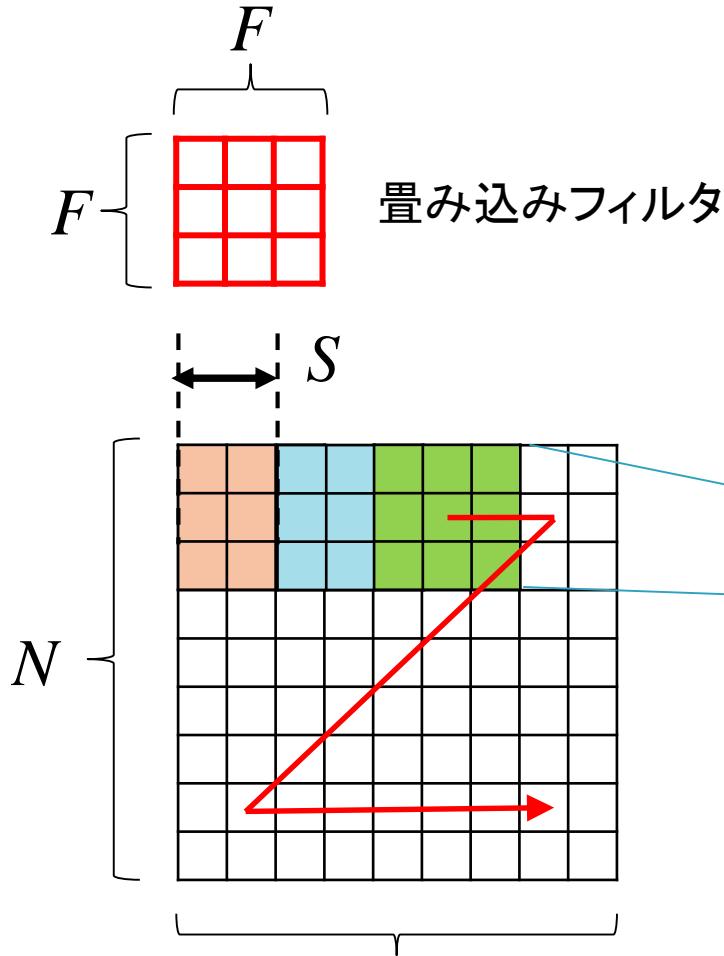
畳み込み層

▶ (まず簡単のため) 入力一層、フィルタ一つの場合

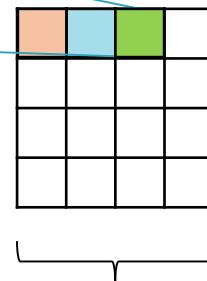


畳み込み層

▶ (まず簡単のため) 入力一層、フィルタ一つの場合



注: 実際は入力層を除き、
 $S=1$ とする場合が多い
(つまり畠み込み層で解像度は落ちない)

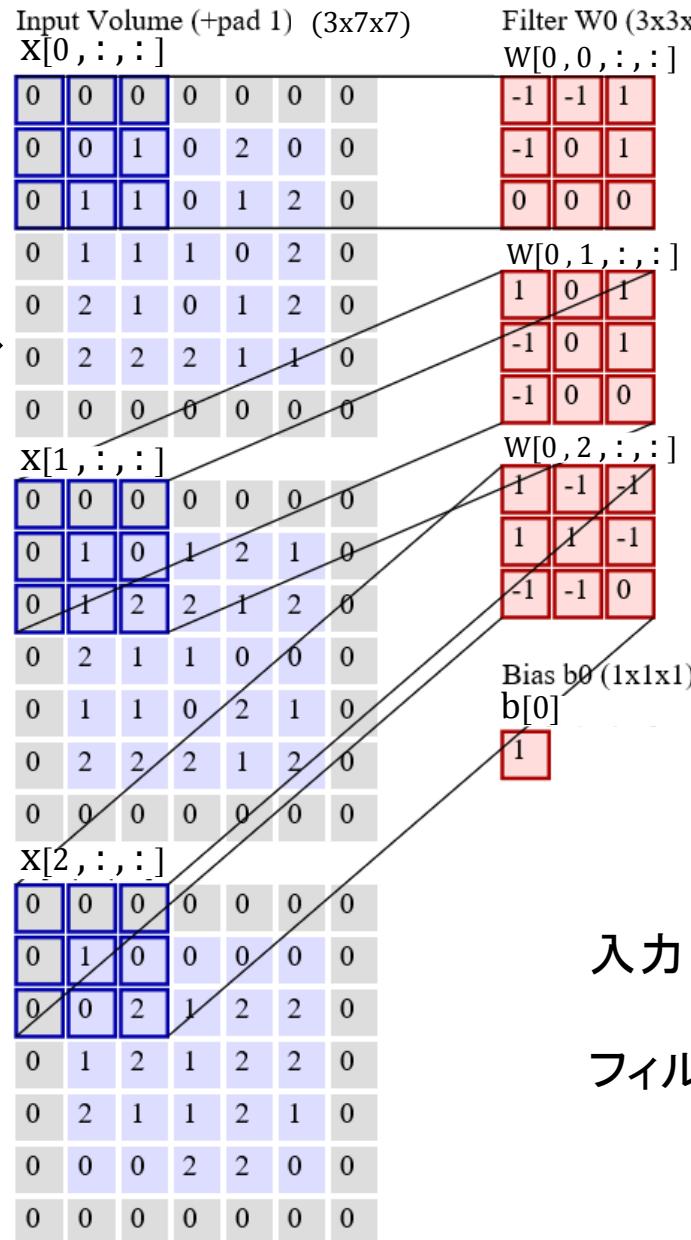
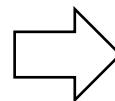


$$(N - F) / S + 1$$

もう少し詳しく

<http://cs231n.github.io/convolutional-networks/> を改変

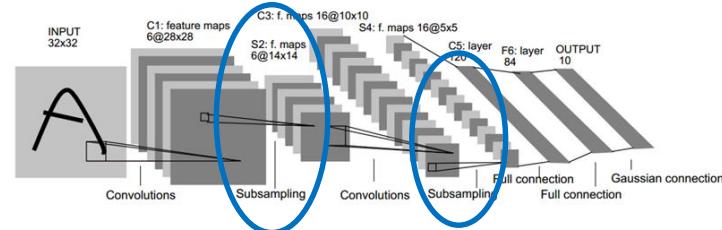
Zero-padding
フィルタがはみ出
す分をゼロ埋め
 $(F - 1)/2$



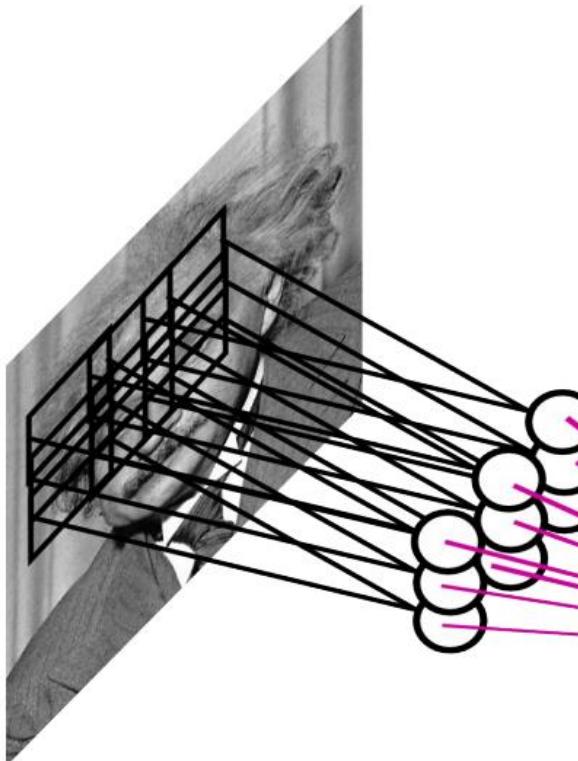
このあと活性化関数
がかかる

入力: チャネル数 3
サイズ N=7 (padding込)
フィルタ: 出力チャネル数 2
サイズ F=3
stride S=2

プーリング層



- ▶ 一定領域内の畳み込みフィルタの反応をまとめる
 - 領域内での平行移動不变性を獲得

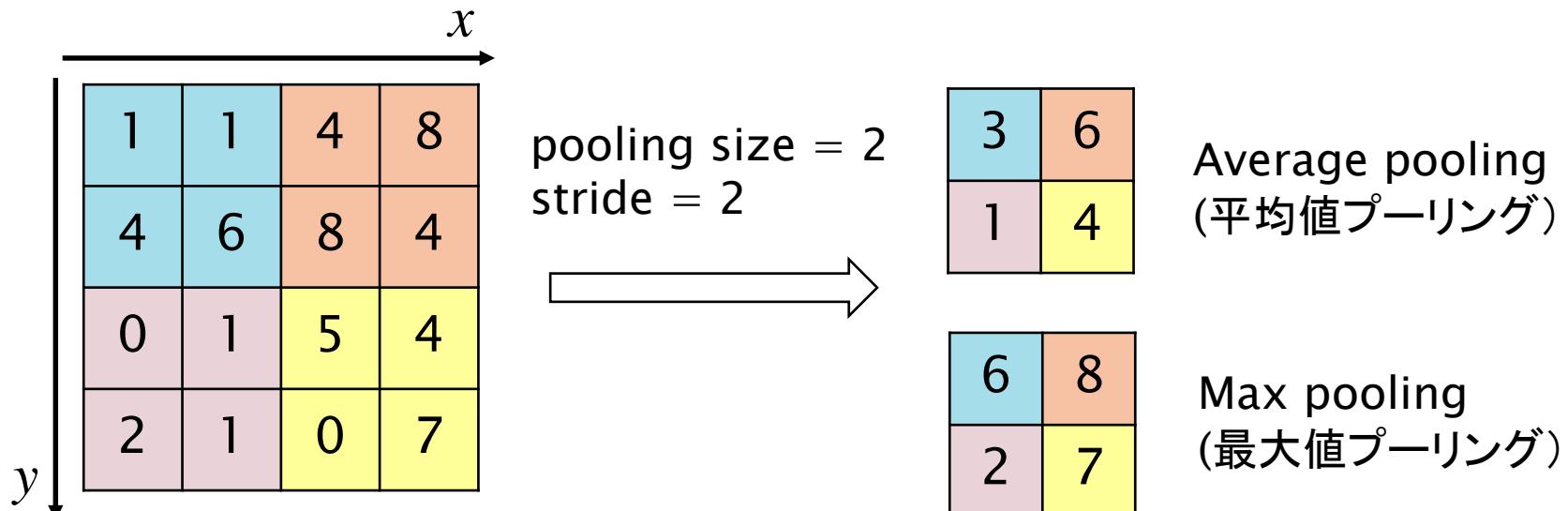


平均値プーリング、
最大値プーリングなど
(古典的には単純なサンプリング)

Source: M. Ranzato, CVPR'14 tutorial slides

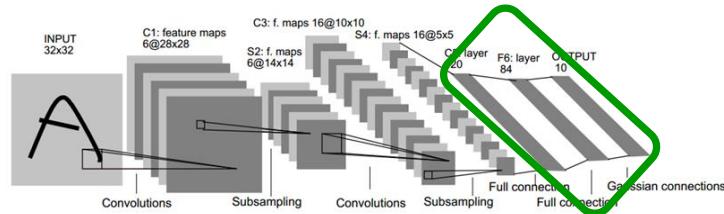
プーリング層

- ▶ Average pooling: 局所領域の平均値をとる
- ▶ Max pooling: 局所領域の最大値をとる

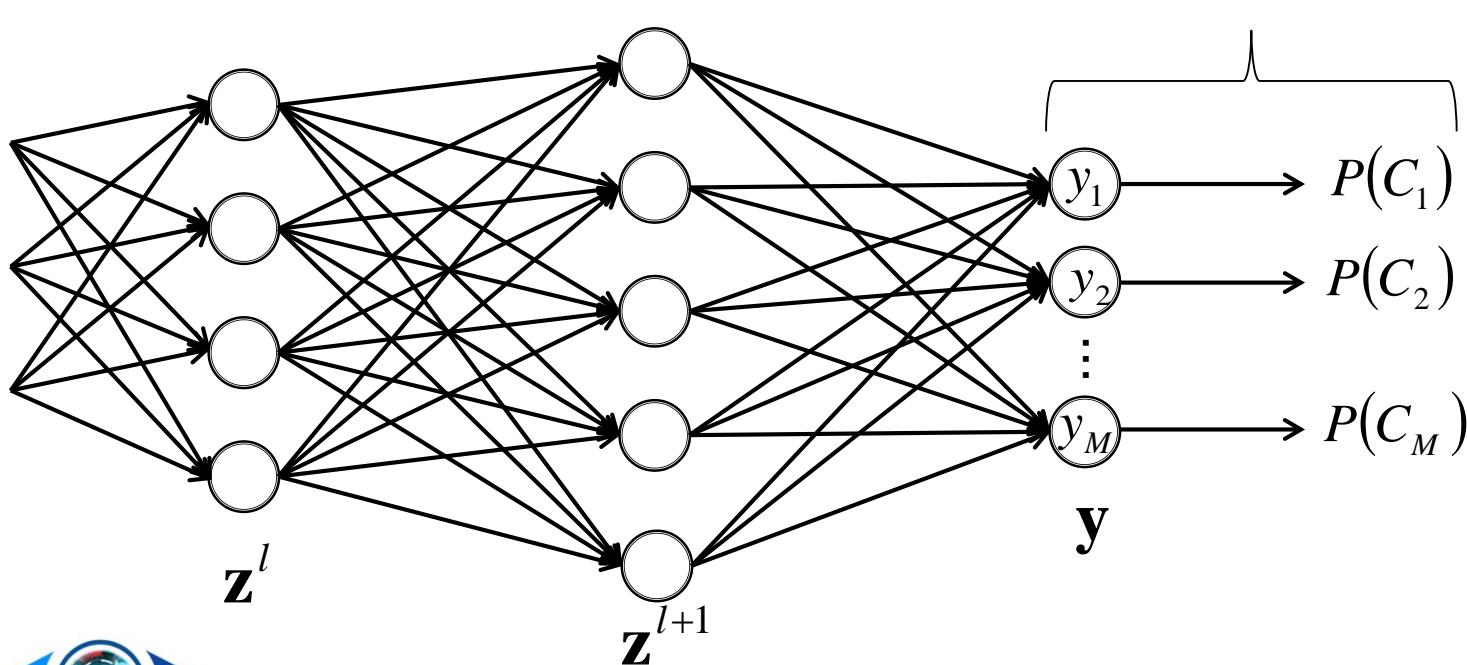


- ▶ 他にも L_p pooling, stochastic pooling などいろいろ

全結合層・出力層



- ▶ 要するにただの多層パーセプトロンです
(第4回講義参照)



Backprop: 置み込み層

▶ 全受容野での誤差を束ねて更新

$z_{k,i}^l = h(u_{k,i}^l)$: k 番目の需要野の i 番目の出力値 (第 l 層)

$\delta_{k,i}^l$: k 番目の需要野の i 番目の誤差値 (第 l 層)

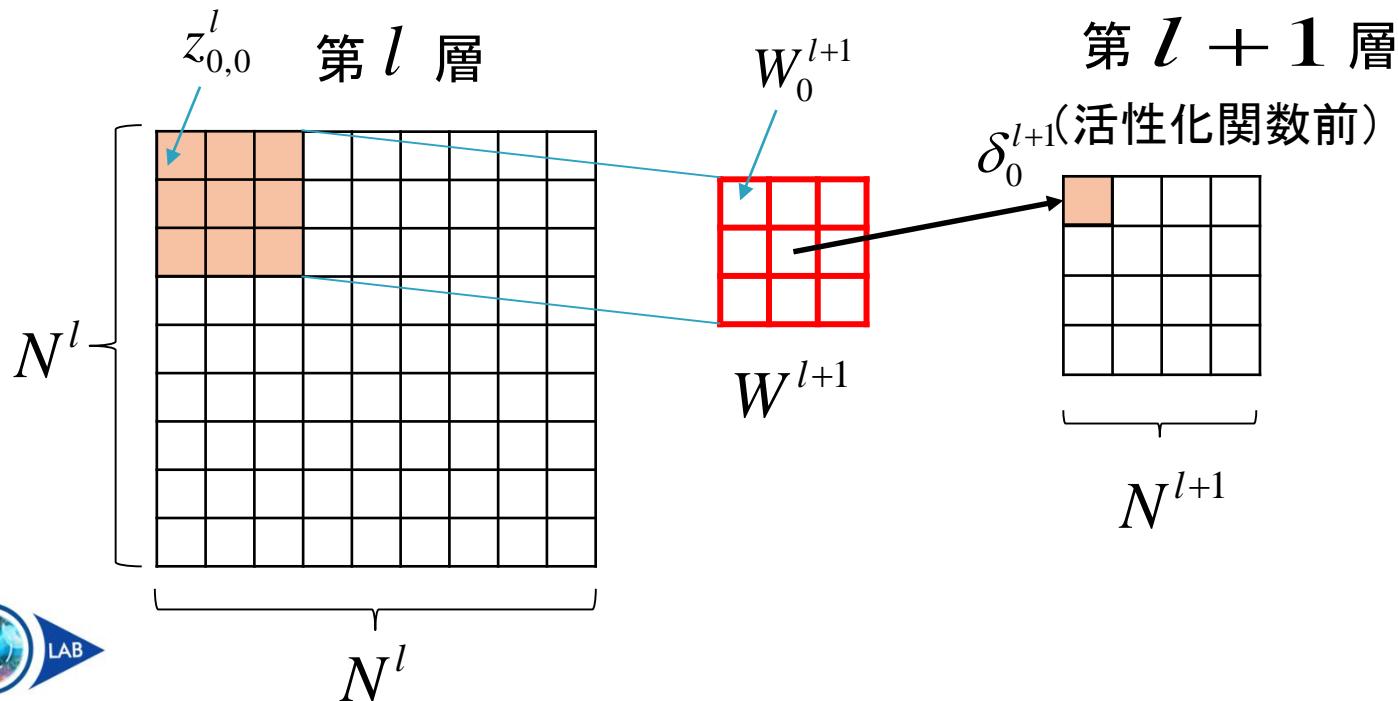
δ_k^{l+1} : 第 $l+1$ 層の対応する場所の誤差

W_i^{l+1} : フィルタの i 番目の係数

$$\frac{\partial J}{\partial W_i^{l+1}} = \sum_{k=0}^{(N^{l+1})^2 - 1} \delta_k^{l+1} z_{k,i}^l$$

$$\delta_{k,i}^l = h'(u_{k,i}^l) \sum_j \delta_j^{l+1} W_{i'}^{l+1}$$

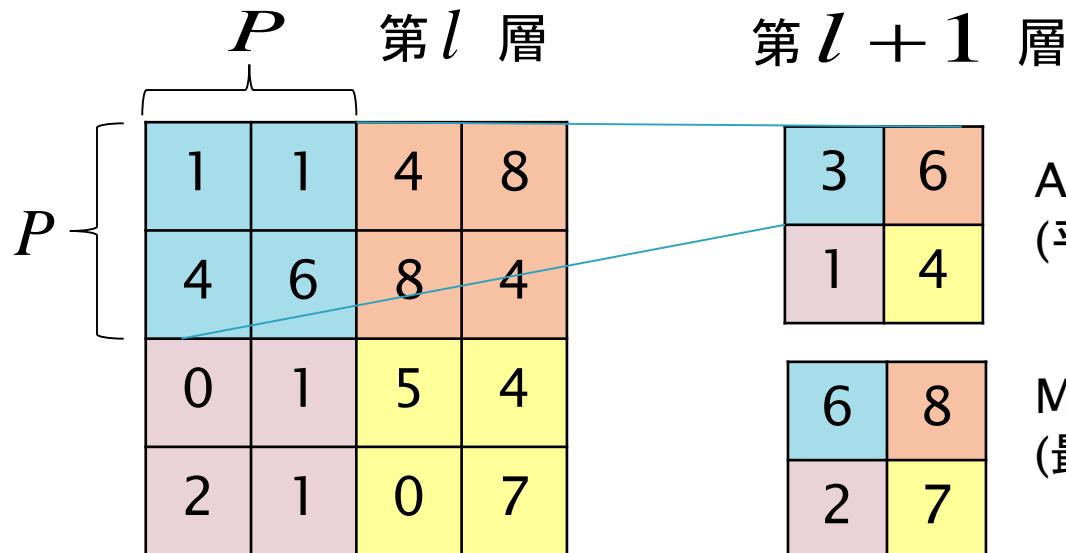

結合がある部分の
フィルタ係数



Backprop: プーリング層

- ▶ Average pooling の場合は簡単
- ▶ Max pooling の場合、feedforward時に選ばれたユニットにのみ誤差が伝播する(覚えておく必要がある)

$$\delta_{k,i}^l = \sum_i^{P^2} \delta_k^{l+1} \underline{W_i^{l+1}}$$



Average pooling
(平均値プーリング) $\delta_{k,i}^l = \delta_k^{l+1} / P^2$

Max pooling
(最大値プーリング)

$$\delta_{k,i}^l = \begin{cases} \delta_k^{k+1} & \text{if } i = \arg \max_j (z_{k,j}^l) \\ 0 & \text{otherwise.} \end{cases}$$

その他

▶ 活性化関数

- 古典的にはシグモイド関数やtanh関数など
- 最近は基本的に全てReLUベース

▶ 最適化

- ミニバッチ法 + モメンタム付きSGD
- 学習率の調整はとても重要(スケジューリング)
- あるいはAdamなどのソルバを使う

▶ Dropout

- 全結合層で行うのが基本
- ただし、最近は全結合層をあまり使わなくなっている(次週)

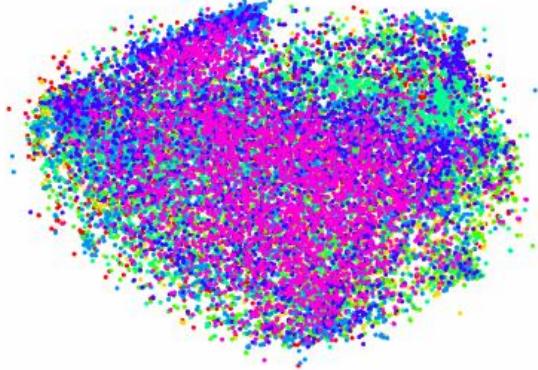


以下捕捉

各層のニューロンの出力

- ▶ 層を上るにつれ、クラスの分離性能が上がる

ILSVRC'12 の
validation data
(色は各クラスを示す)



(c) DeCAF₁

第1層



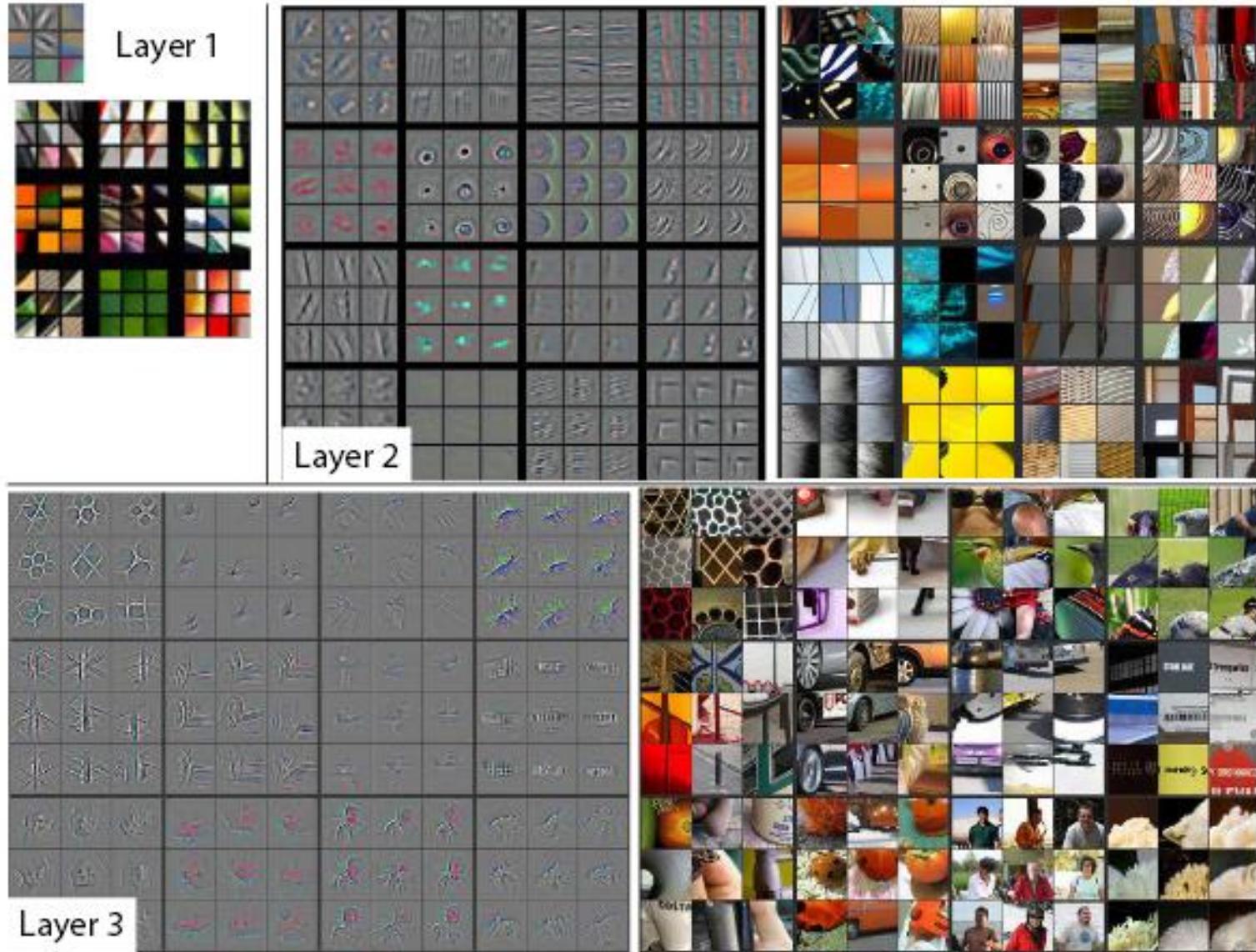
(d) DeCAF₆

第6層

J. Donahue et al., “DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition”, In Proc. ICML, 2014.

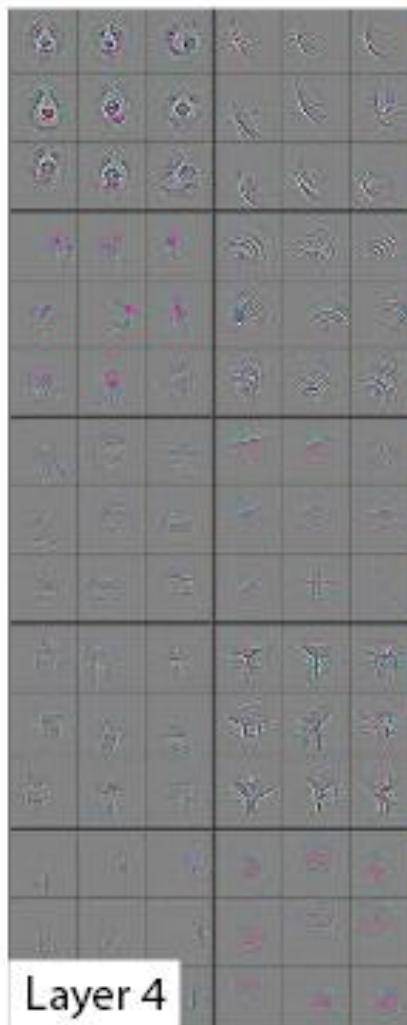
中間層の可視化

Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks", In Proc. ECCV, 2014.



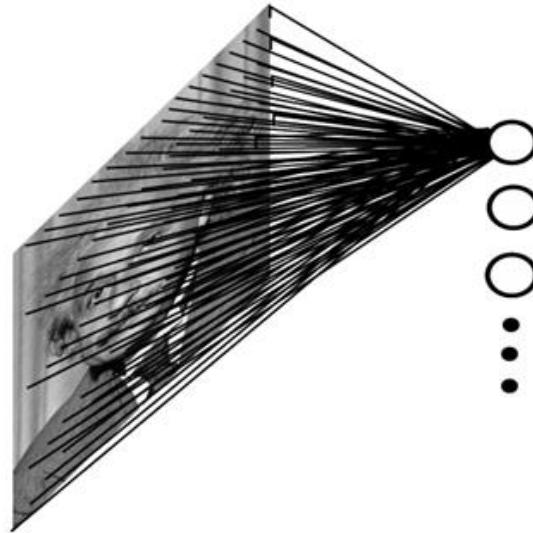
中間層の可視化

Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks", In Proc. ECCV, 2014.



ConvNet以外のアーキテクチャはどうか？

- ▶ 全結合ネットワーク
 - 極めて多くのパラメータ
 - 最適化が困難
 - 収束まで時間がかかる
 - そもそもメモリにのらない



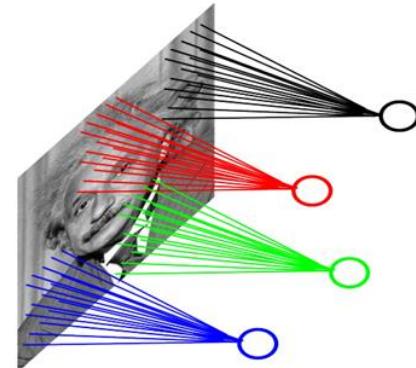
Source: M. Ranzato, CVPR'14 tutorial slides

MNISTデータセット(28x28ピクセル)のような小さい画像を用いて古くから研究されているが、今のところConvNetには及ばない

ConvNet以外のアーキテクチャはどうか？

▶ 局所結合ネットワーク

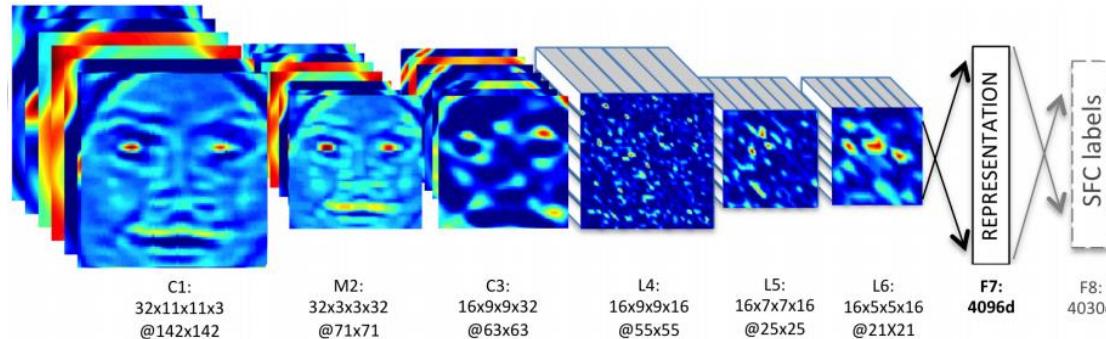
- 構造はConvNetと同じだが、フィルタのパラメータに場所ごとで異なる
- つまり、平行移動不変性がない



Source: M. Ranzato, CVPR'14 tutorial slides

▶ 入力画像の正確なアラインメントが前提となっている場合、state-of-the-art を達成した場合もある

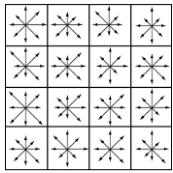
- DeepFace [Taigman et al., CVPR' 14]



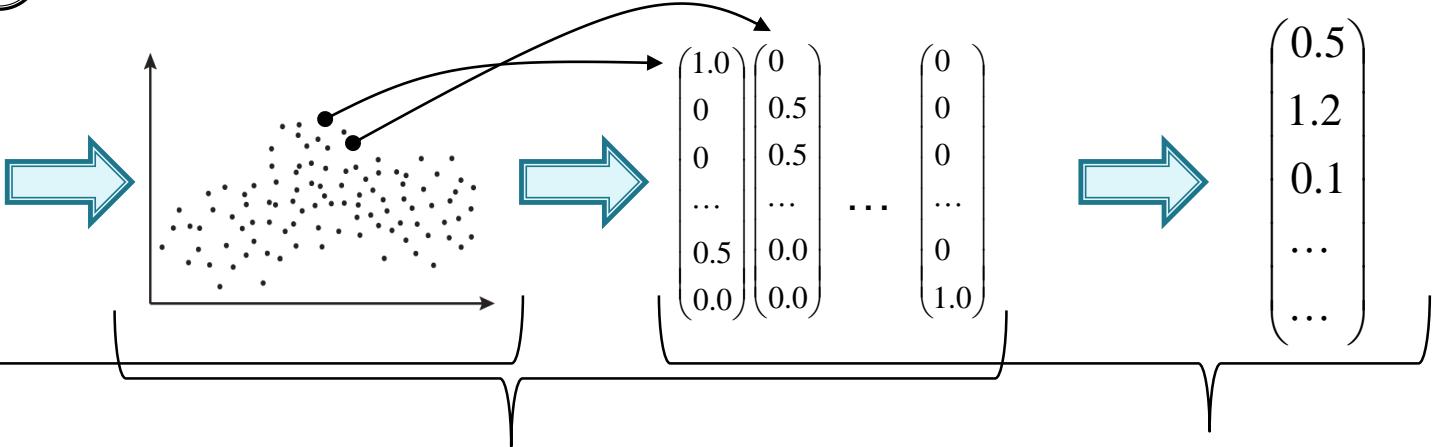
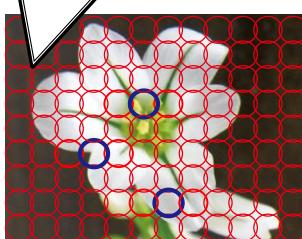
▶ 一般的な画像認識ではまだConvNetに劣る

深層学習以前の画像特徴抽出の枠組

- ▶ 画像中の局所特徴の分布(統計情報)を表現する大域的特徴ベクトルを抽出



e.g.
SIFT記述子



1. 局所特徴抽出

- SIFT, SURF, HOG, etc.
- Dense sampling
(回転、スケールの正規化なし)

2. エンコーディング

- ベクトル量子化
- 多項式特徴(要素積)

3. プーリング

- 最大値プーリング
- 平均値プーリング

Bag-of-Visual-Words (BoVW) [Csurka et al. 2004]

- ▶ ベクトル量子化により局所特徴のヒストグラムを作成

